

# ANWAR SHAIKH

DATA ENGINEER

in Anwar Shaikh



(+91)-8237685256



anwarshaikh078@gmail.com



India

## **ABOUT**

Data engineering professional with 3 years of expertise in technology and consulting. Passionate in assisting organizations in making data-driven choices through the use of powerful ETL pipelines, cloud engineering, analytics, and creative visualization. A firm believer in building and developing quick, trustworthy, and adaptive data environments.

## **EDUCATIONAL HISTORY**

### **BE. COMPUTER SCIENCE**

MGM's JNEC, Aurangabad Completed - 2019

#### **SENIOR SECONDARY**

Maulana Azad CollegeAurangabad Completed - 2015

#### SECONDARY.

St. Xavier's, Aurangabad Completed - 2013

## **PROJECTS**

- **Jnec Agnitio** An Android app for the annual technical fest that allows you to view descriptions of all technical events and book them. It received over 500 downloads in the first seven days on the Google Play store.
- File Explorer with Cloud Storage A file management desktop application. You can save files, images, videos, and other media directly to your Google Drive without having to open it. We used the Google Drive API and some C# system libraries for this project.

## **DEEP-LEARNING PROJECTS**

- <u>Digit Recognizer using OpenCV and CNN</u>
- LSTM Sentiment Analysis
- Art Generation using Neural Style Transfer.
- Emotion Recognition
- Invisibility Cloak.

## **MEDIUM ARTICLES**

- Real-Time in GCP.
- Workload Management using Bigguery **Reservation Slots**

## AREAS OF INTEREST

- System and Database Designs.
- Machine Learning and Al.
- Business and Data Analytics.

# **PROFESSIONAL HISTORY**

## DATAMETICA SOLUTIONS PVT LTD.

July 2019 - Present

## **ADT SECURITY SERVICES | DATA ENGINEER 2**

- Worked on Data Ingestion using Dataflow.
- Worked on improving the performance of **Airflow** Dags. One DAG used to take 13 hours to complete, but it now only takes 3.5 hours.
- DAGs were created to load data from Google Cloud Storage into Bigguery and Cloud SQL.
- Using Google Cloud Composer, I created DAGs (Directed Acyclic Graphs) and scheduled them in Tidal.
- Using **SQL**, I created multiple stored procedures and achieved scd-1 and scd-2 in the data warehouse.

#### **GAMESTOP | DATA ENGINEER 1**

- I spent a lot of time working on the streaming data pipeline. To accomplish this, we used Google Cloud Pub/Sub, Cloud Functions, and Google Cloud Bigguery.
- Audit Framework was developed in order to implement change data capture (CDC).
- Hadoop Native jobs were converted to GCP equivalents.
- Developed Pyspark (Spark) scripts to load data into Bigguery using **Dataproc**.
- I worked on File Ingestions (i.e. CSV, txt) in order to load them into Google Bigguery.
- I worked on extracting data from Informatica files.
- Worked on several Python scripts to pre-process data, merge, and so on.

## **SKILLS**

- **GCP Services** GCS, Cloud Functions, PubSub, Composer, Dataflow, Dataproc, Bigquery, Cloud SQL, Cloud Spanner, Compute Engine, VPC.
- AWS Services RDS, Redshift, Lambda, EC2, S3.
- Databases Teradata, Mongo DB, My SQL, Netezza.
- **Programming** Python, Java, Node.js, Unix, C++, PySpark.
- **BigData** Hadoop, MapReduce, Apache Spark, Apache Kafka, Apache Airflow, Apache Beam.
- Others Git, Gitlab, Jira, Terraform, Docker, Kubernetes, Looker.

## **ACHIEVEMENTS**

- Hulk award, Rising Star award from Datametica.
- Principal Nominee for the year 2018-2019.
- HOD Nominee CSE Department for 2017-2018.
- Qualified many international coding competitions i.e Facebook HackerCup, Google CodeJam.
- Participated in Abu-Robocon-2018.

# **EC MOBILITY** - SOFTWARE DEVELOPMENT INTERN

June 2018 - July 2018

- Created a reporting tool for a company with over 100 employees.
- The tool was built on the Oracle Apex Framework and includes inventory management methods.
- Data export from CSV to Relational DB Tables, as well as various SQL operations for proper analysis and insights.

## **CERTIFICATIONS**

- Apache Spark 3 Real-Time Stream Processing using Python.
- Apache Spark 3 Spark programming in Python.
- ML for Business Professionals.
- Data Science Math Skills.
- Structuring Machine Learning Projects.
- Convolutional Neural Network.
- Sequence Model.

## **VOLUNTEER WORK**

- Student Co-Ordinator in annual cultural fest i.e Razzmatazz-2018-2019.
- Organized many events in annual cultural fests over the 3 years.
- Worked with training and placement cell for 2 years.